Real Dataset Analysis

```r
# Load preprocessed dataset
bladder_comp_adj <- readRDS(here("paper", "data", "bladder_comp_adj.rds"))

# Create a stratified 50/50 split
set.seed(123)
split <- initial_split(bladder_comp_adj, prop = 0.5, strata = event)

# Create training and testing data frames
train <- training(split)
test  <- testing(split)

# Verify the proportions
table(train$event) / nrow(train)
```

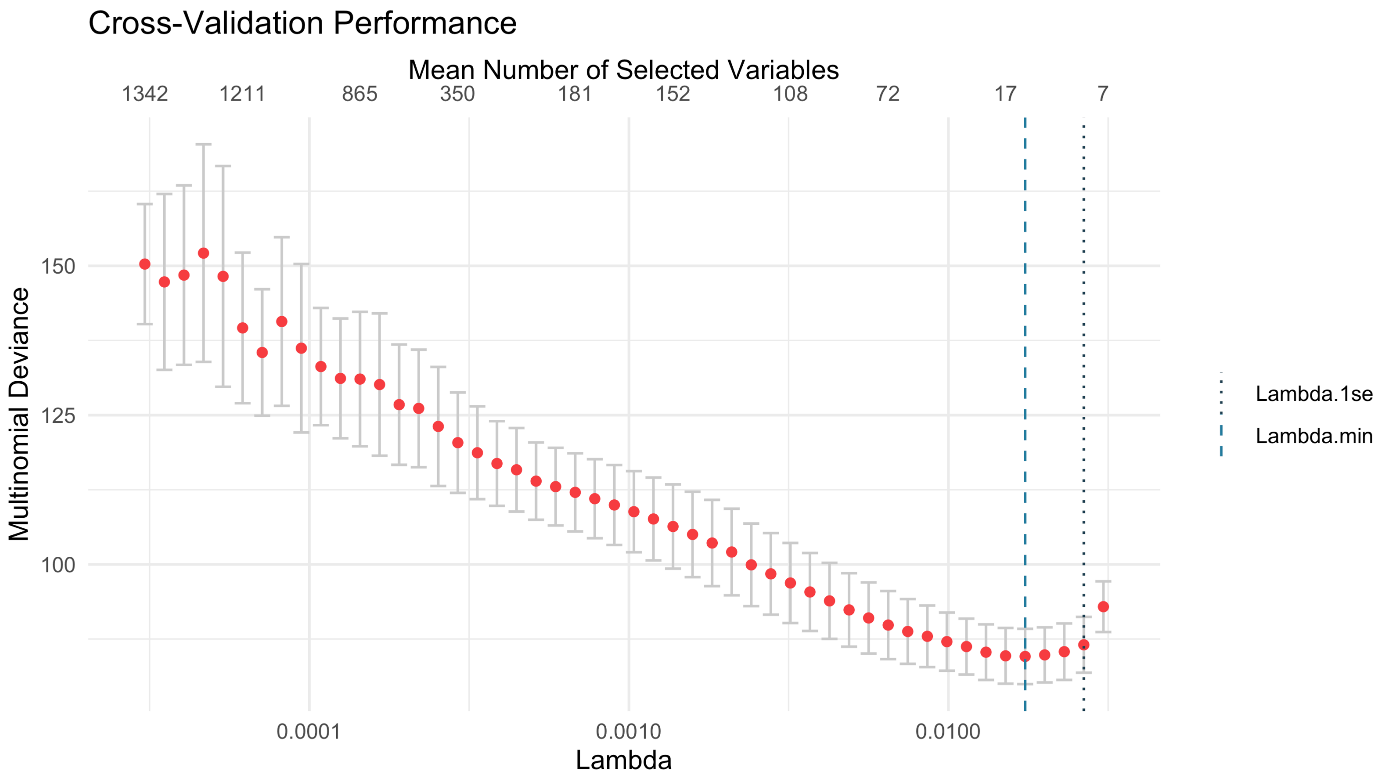
```
#> 
#>         0         1         2 
#> 0.7466667 0.0800000 0.1733333
```

```r
table(test$event) / nrow(test)
```

```
#> 
#>          0          1          2 
#> 0.76158940 0.05298013 0.18543046
```

1 cbSCRIP

```r
# cbSCRIP cross validation

## Check if file has been run before.
file_save <- here("paper", "results",
             glue("cv_multinom_enet.rds"))

if (!file.exists(file_save)) {

  set.seed(123)
  # Run CV
  tictoc::tic()
  cv_multinom_enet <- cv_cbSCRIP(
     Surv(time, event) ~ .,
     cbind(train[,-(2:7), , drop = FALSE],
         train[,2:7, , drop = FALSE]),
     n_unpenalized = 7,
     lr_adj = 50,
     maxit = 200,
     fit_fun = cbSCRIP::MNlogisticSAGA_Native)
  tictoc::toc()

  write_rds(cv_multinom_enet,
        here("paper",
           "results",
           glue("cv_multinom_enet.rds")))
}

cv_multinom_enet <- readRDS(file_save)

# CV plot
plot(cv_multinom_enet)
```



```
# Model results
cv_multinom_enet
```

```
#> --- Cross-Validated Case-Base Competing Risks Model ---
#>
#> Call:
#> cv_cbSCRIP(formula = Surv(time, event) ~ ., data = cbind(train[,
#>   -(2:7), , drop = FALSE], train[, 2:7, , drop = FALSE]), n_unpenalized = 7,
#>    lr_adj = 50, maxit = 200, fit_fun = cbSCRIP::MNlogisticSAGA_Native)
#>
#> Performed 5-fold cross-validation over 50 lambda values.
#>
#> Optimal Lambda Values:
#>   Lambda with minimum deviance (lambda.min): 0.0174
#>   Largest lambda within 1 SE of min (lambda.1se): 0.0265
#>
#> The final model (fit.min) was fit using lambda.min and has 17 non-zero coefficients.
```

```
cv_multinom_enet$fit.min
```

```
#> -- Case-Base Competing Risks Model (cbSCRIP) -----------------------------------
#>
#> i Lambda: 0.0174
#> i Alpha (for elastic-net): 0.5
#> -- Selected Coefficients -----------------------------------------------------
--
#> Found 17 non-zero coefficients:
#>
#>  Variable    Cause Coefficient
#>  age         1      0.01505
#>  female      1      0.32733
#>  treatment   1     -0.12847
#>  gradeHIGH   1      0.44448
#>  stageT1     1     -0.65352
#>  log(time)   1     -0.47047
#>  (Intercept) 1     -5.71578
#>  seq34       2      0.13023
#>  seq78       2      0.11604
#>  seq982      2      0.01138
#>
#> ... and 7 more non-zero coefficients.
```

```
cv_multinom_enet$fit.min$convergence_pass
```

```
#> [1] 129
```

```
# models w/ lambda min and 1SE
multinom_enet <- cbSCRIP(
  cb_data = cv_multinom_enet$cb_data,
  n_unpenalized = 7,
  lr_adj = 50,
  maxit = 200,
  fit_fun = cbSCRIP::MNlogisticSAGA_Native,
  warm_start = F,
  lambda = c(cv_multinom_enet$lambda.min,
         cv_multinom_enet$lambda.1se))

# Print selected coefficients
```

```r
coefs_min <- multinom_enet$coefficients[[which(multinom_enet$lambdagrid == cv_multinom_ene

coefs_min[rowSums(abs(coefs_min)) > 1e-8, ]
```

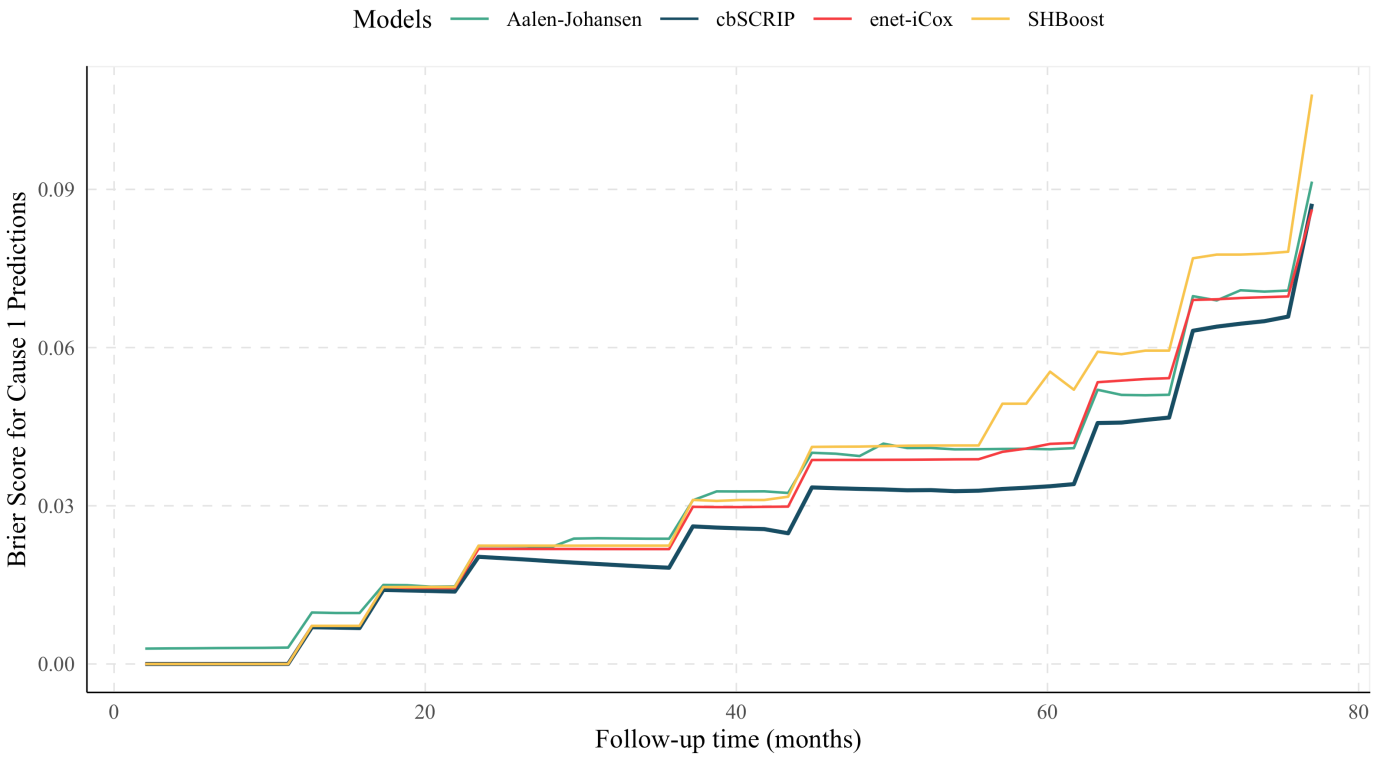
```
#>                     [,1]        [,2]
#> seq34        -0.38201105  0.90498100
#> seq78        -0.30626439  0.23387158
#> seq982       -0.76867996  0.32870105
#> age           0.04419912  0.03612058
#> female        1.10200989 -0.51480246
#> treatment    -0.81356794 -0.47415546
#> gradeHIGH     0.94668359  1.56795064
#> stageT1      -0.81515012 -0.55449334
#> log(time)     1.47734964 -0.24282754
#> (Intercept) -15.49276762 -8.00781670
```

```r
coefs_1se <- multinom_enet$coefficients[[which(multinom_enet$lambdagrid == cv_multinom_ene

coefs_1se[rowSums(abs(coefs_1se)) > 1e-8, ]
```

```
#>                     [,1]        [,2]
#> age           0.05005823  0.03631520
#> female        0.85876052 -0.33633975
#> treatment    -0.76419507 -0.55319168
#> gradeHIGH     1.00586257  1.65861676
#> stageT1      -1.21482854 -0.08147097
#> log(time)     1.43310284 -0.36186966
#> (Intercept) -15.17921903 -8.02930712
```

```r
# Brier Score plot
# Set time points
time_points <- sort(unique(test$time))

# Restrict test time-points outside the training
tmax <- max(train$time[train$event != 0], na.rm = TRUE)

valid_event_times <- sort(unique(test$time[test$event != 0]))
valid_event_times <- valid_event_times[valid_event_times <= tmax]

time_points <- seq(from = min(valid_event_times),
             to = max(valid_event_times),
             length.out = 50)

## Casebase
casebase_fit <- casebase::fitSmoothHazard(
  formula = event ~ .,
  data = train[,1:7, , drop = FALSE],
  time = "time",
  ratio = 50
```

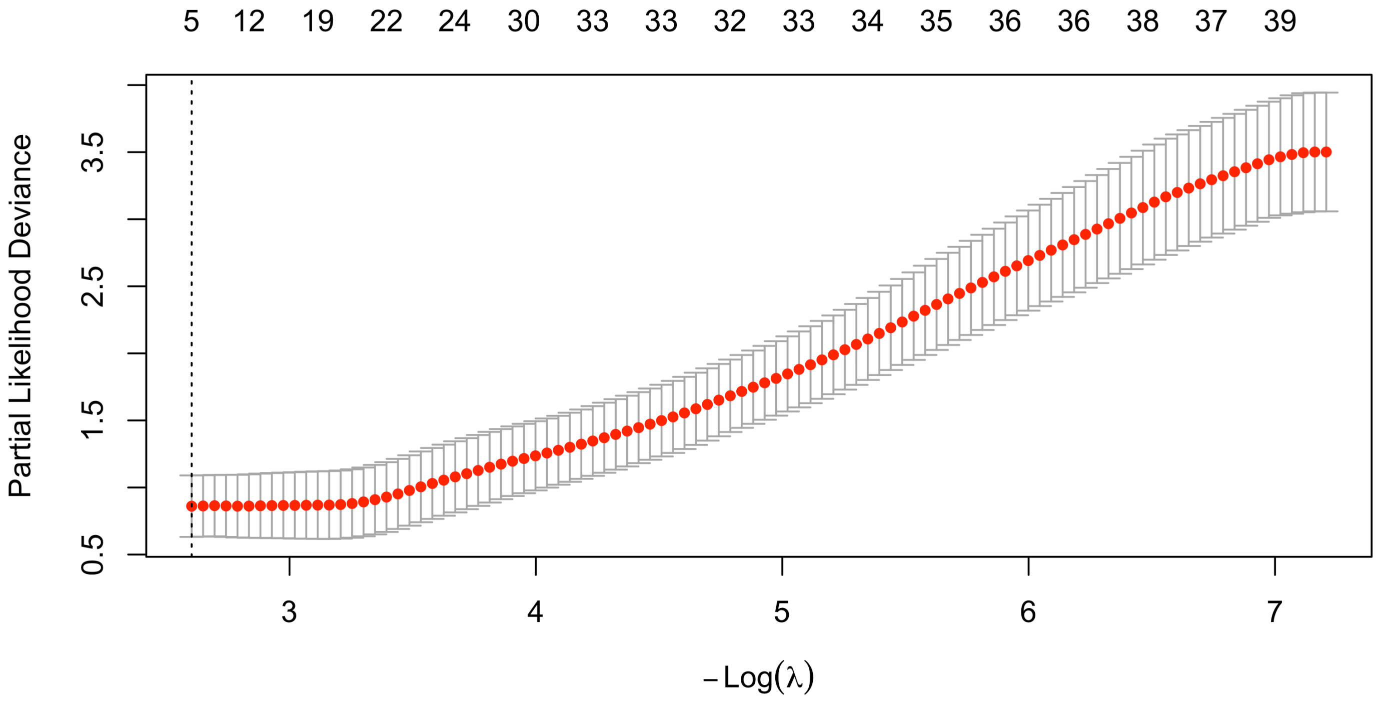
```r
)

## Aalen-Johnson
aj_fit <- prodlim(Hist(time,event)~1, data= train)

## iCoxBoost
optim.res <- iCoxBoost(Hist(time, event)~., data = train, cause = 1, cv = TRUE)

icox_fit <-  iCoxBoost(Hist(time, event)~., data = train, cause = 1,
             stepno = optim.res$cv.res$optimal.step)

# Cause-Specific
icr_fit = two.i.CSlassos(data = train, nlambda = 100,
              var_time = "time",
               var_status = "event")

all_models <- list(
  "cbSCRIP 1SE" = multinom_enet$refitted_models[[which(
    multinom_enet$lambdagrid == cv_multinom_enet$lambda.1se
    )]],
  "cbSCRIP" = multinom_enet$refitted_models[[which(
    multinom_enet$lambdagrid == cv_multinom_enet$lambda.min
    )]],
  "Casebase" = casebase_fit,
  "Aalen-Johansen" = aj_fit,
  "SHBoost" = icox_fit,
  "enet-iCox" = icr_fit
)

# Estimate Brier Scores
all_scores_event_1 <- Score(
  all_models,
  formula = Hist(time, event) ~ 1,
  data = test,
  times = time_points,
  summary = "ibs",
  se.fit = FALSE,
  metrics = "Brier",
  cause = 1
)

all_scores_event_2 <- Score(
  all_models,
  formula = Hist(time, event) ~ 1,
  data = test,
  times = time_points,
  summary = "ibs",
  se.fit = FALSE,
  metrics = "Brier",
  cause = 2
)

data_brier <- all_scores_event_1$Brier$score %>%
  filter(model != "Null model") %>%
  mutate(model = as.character(model))

# Plot Brier Score

cpl_palette <- c("cbSCRIP" = "#184d63",
           "cbSCRIP 1SE" = "#277DA1",
           "Casebase" = "#16A1CD",
           "Aalen-Johansen" = "#43AA8B",
           "SHBoost" = "#F9C74F",
           "enet-iCox" = "#f94144")

(plot_ar_briers <- data_brier |>
    filter(!str_detect(model, "cbSCRIP 1SE|Casebase")) |>
 ggplot(aes(times, Brier, colour = model,
         linewidth = ifelse(str_detect(model, "cbSCRIP"), 0.5, 0.1))) +
  geom_line() +
  scale_colour_manual(values = cpl_palette) +
  scale_linewidth_continuous(range = c(0.5, 0.8),
                    guide = "none") +
  labs(x = "Follow-up time (months)",
      color = "Models",
      y = "Brier Score for Cause 1 Predictions"))
```



```
ggsave(here("paper",
       "figs",
       "bladder_comp_brier_score.png"),
       plot_ar_briers,
       width = 20, height = 12, units = "cm", dpi = 300)
```

2 Cause-specific Cox models with LASSO

```
y <- Surv(time = train$time,
      event = as.numeric(train$event == 1))

x <- model.matrix(event ~ . -time,
           data = train)

p.fac <- rep(1, ncol(x))
p.fac[1:6] <- 0

set.seed(1234)
cox_lasso_mod <- cv.glmnet(x = x, y = y, family = "cox",
                # family = "binomial",
                penalty.factor = p.fac,
                nfolds = 5,
                alpha = 1,
                thresh = 1e-9,
                maxit = 1e9)

plot(cox_lasso_mod)
```



```
cc_lasso_min <- coef(cox_lasso_mod, s = cox_lasso_mod$lambda.min)

select_vars_lasso <- cc_lasso_min@Dimnames[[1]][-1][cc_lasso_min@i]

selected_coefs_lasso <- cc_lasso_min@x

names(selected_coefs_lasso) <- select_vars_lasso

selected_coefs_lasso
```

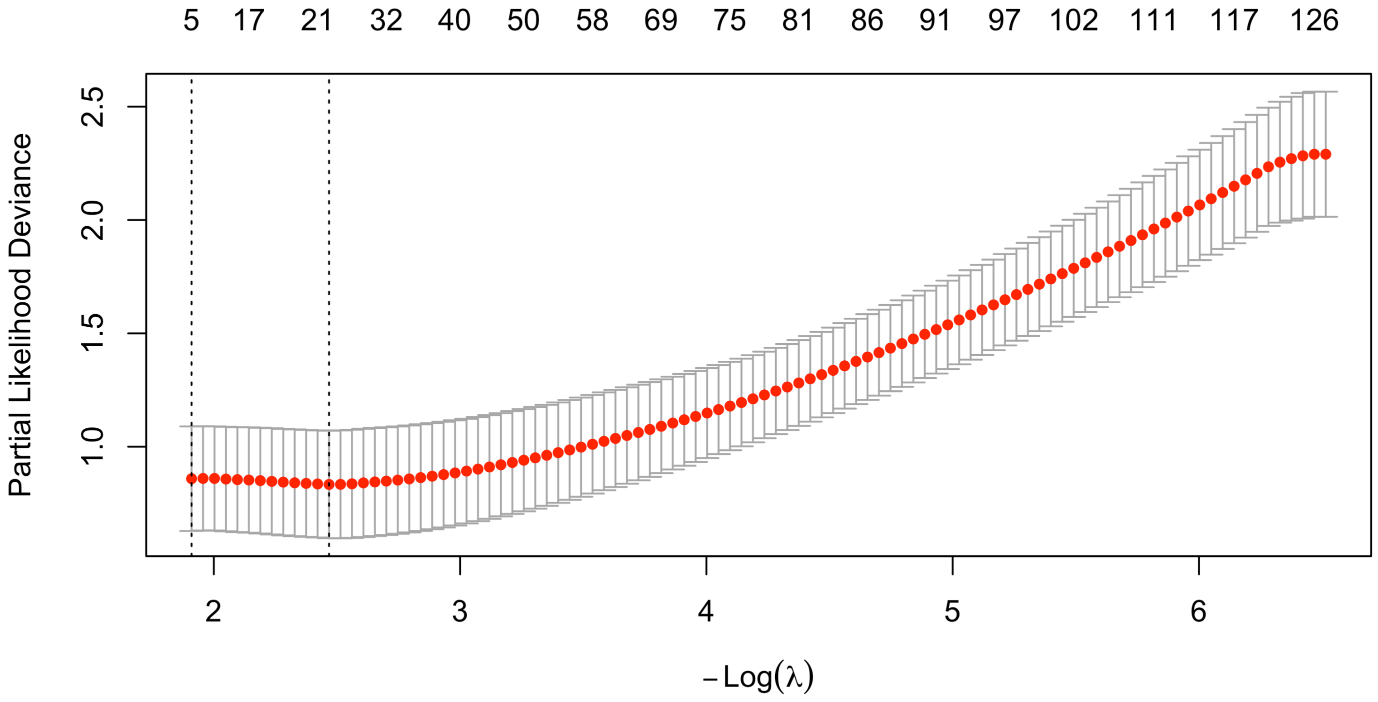
```
#>         age      female   treatment   gradeHIGH     stageT1
#>  0.04611944  0.77433108 -0.92522786  0.85713859 -1.00346477
```

```
length(selected_coefs_lasso)
```

```
#> [1] 5
```

3 Cause-specific Cox models with enet

```
p.fac <- rep(1, ncol(x))
p.fac[1:6] <- 0

set.seed(1234)
cox_enet_mod <- cv.glmnet(x = x, y = y, family = "cox",
                # family = "binomial",
                penalty.factor = p.fac,
                nfolds = 5,
                alpha = 0.5,
                thresh = 1e-9,
                maxit = 1e9)

plot(cox_enet_mod)
```



```
cc_enet_min <- coef(cox_enet_mod, s = cox_enet_mod$lambda.min)

select_vars_enet <- cc_enet_min@Dimnames[[1]][-1][cc_enet_min@i]

selected_coefs_enet <- cc_enet_min@x

names(selected_coefs_enet) <- select_vars_enet

selected_coefs_enet
```

```
#>          age       female    treatment    gradeHIGH      stageT1      seq1036 
#>  0.049055788  0.853859222 -1.322648822  1.081615276 -1.058595729 -0.330860959 
#>      seq1055      seq1070      seq1101      seq1103      seq1115      seq1118 
#> -0.163720347 -0.002244496 -0.023715439 -0.282596236 -0.123230918 -0.236495369 
#>      seq1133       seq121      seq1299      seq1335    seq1384_2       seq281 
#> -0.102276503 -0.200942973  0.149561506 -0.082274307 -0.062000988  0.235680777 
#>       seq353       seq580       seq679       seq739       seq761 
#> -0.142889838  0.191734699 -0.134233346  0.007147549 -0.120710915
```

```
length(selected_coefs_enet)
```

```
#> [1] 23
```